IN THE UNITED STATES PATENT AND TRADE MARK OFFICE

VERIFICATION OF TRANSLATION

I, Dr Lesley Margaret Stone, B.Sc., Ph.D., D.M.S., of Silver Creek, Lhergy Cripperty, Union Mills, Isle of Man, IM4 4NJ, British Isles, do hereby declare that I am conversant with the English and French languages and that I am a competent translator thereof;

I verify that the attached English translation of International Patent Application Number PCT/FR02/00951 filed on 18th March 2002 is a true and correct translation made by me of the attached document in the French language;

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that wilful false statements and the like so made are punishable by fine or imprisonment or both under Section 1001 of Title 18 of the United States code and that such wilful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 12th June 2004 Lesley Stone

NUCLEIC ACID MOLECULES ENCODING A DEXTRANSUCRASE CATALYZING THE SYNTHESIS OF DEXTRAN CARRYING ALPHA-1, 2-OSIDIC TYPE LINKAGES

The present invention relates to the field of glycotechnology, more particularly to the synthesis of oligosaccharides or oligosides with a prebiotic, therapeutic or diagnostic effect.

The present invention pertains to nucleic acid molecules encoding an enzyme having a glycosyltransferase activity catalyzing the synthesis of dextrans or oligosides carrying α(1→2) osidic type linkages.

The invention also pertains to enzymes synthesized by the nucleic acids of the invention, and to their expression systems in prokaryotic or eukaryotic cells. Finally, they pertain to the use of said enzymes in the production of oligosaccharides in foodstuffs, or as an active principle in therapeutic and/or cosmetic products.

Oligosides and heterooligosides act as recognition and effector signals in both animals and plants (as oligosaccharines) by specifically binding to lectins, glycosyltransferases, glycosidases, adhesion molecules etc. The antigenic determinants of blood groups are osidics and our defense against many pathogenic bacteria is directed against osidic structures of the bacterial envelope. Further, one of the major reasons for xenograft rejection is the existence of osidic structures belonging to each species. Such properties, and the knowledge acquired in recent years regarding glycosyltransferases and lectins, contribute to making certain oligosides the candidates of choice for therapeutic or prophylactic treatment of disorders linked to the microbiological equilibrium of various organs such as the intestine or skin. As an example, oligosides constitute an interesting alternative to the use of micro-organisms and antibiotics in regulating the composition of intestinal flora (prebiotic effect). Certain oligosides can be considered to be "soluble fiber" when they are not metabolized by human and animal digestive enzymes; on reaching the colon, they interact with the microbial flora and specifically affect the growth and adhesion of certain species. If they are incorporated into food in low doses (less than 1%), certain osidic molecules improve health and stimulate weight gain in animals.

A review of different glycosyltransferases, their structure and their activity, has been carried out by Vincent Monchois et al (1). Briefly:

a) it appears that the structure of the glycosyltransferases and/or dextransucrases studied is highly conserved and is constituted, starting from the amino part of the protein, by a signal sequence, a variable domain, a catalytic domain and a glucan binding domain.

b) glucooligosides (GOS) can be synthesized by glycosyltransferases such as dextransucrases from cheaper substrates such as saccharose and in the presence of a glucose accepting sugar. Other substrates such as α-D-fluoroglucose, para-nitrophenyl-α-D-glucopyranoside, α-D-glucopyranoside-α-D-sorbofuranoside or 4-O-α-D-galactopyranosylsucrose can also be used.

Starting from the substrate, such enzymes catalyze the transfer of glucose units onto acceptor molecules. In the presence of a glucose acceptor such as maltose or isomaltose, glycosyltransferases catalyze the synthesis of low molecular weight oligosaccharides primarily comprising chains with 3 to 7 glucoses, in accordance with the reaction:

Saccharose + acceptor → glucosylated acceptor + fructose

In such cases, enzymes generally have a specificity for the synthesis of osidic bonds in accordance with that forming the donor polymer.

In contrast, in the absence of an acceptor, the enzyme synthesizes high molecular weight dextran type glucans by successive transfer of α-D-glucopyranosyl units from saccharose in accordance with the reaction:

n-saccharose → $(glucose)_n$ + n-fructose

c) The structures and function of glucans or oligosides synthesized by glycosyltransferases depends on the producing bacterial strain.

Throughout the present text, the generic term "glycosyltransferases" is used to designate the different enzymes capable of catalyzing the synthesis of glucose polymers from saccharose. They are generally produced by bacterial strains of the Leuconostoc, Lactococcus, Streptococcus

or Neisseria type. The size and structure of the glucans produced depends on the producing strain.

The glucose units are coupled by α(1→6) osidic bonds as in dextran, by α(1→3) bonds as in the case of mutane, or by alternations of the two types (alternane).

Similarly, the existence and nature of the linkages, their length and position varies depending on the origin of the producing strain.

Glycosyltransferases producing glucans or GOSs containing at least 50% α(1→6) bonds are termed dextransucrases. GOSs synthesized by said enzymes may carry α(1→2), α(1→3) and/or α(1→4) linkages. Said dextransucrases are produced by Leuconostoc mesenteroides type bacteria.

d) The dextransucrase from L. mesenteroides NRRL B-1299 can produce a highly branched dextran the majority of linkages of which are of the α(1→2) type. Used in the presence of saccharose and maltose, a glucose acceptor molecule, it results in the formation of GOS some of which have a α(1→2) bond at their non-reducing end and others of which have α(1→2) linkages on intermediate residues between the ends. For this reason, they resist degradation by enzymes (hydrolases) of the upper digestive tract in man and animals, and are only degraded by bacterial genuses that are capable of fermenting, such as Bacteroides and Bifidobacterium, considered to be beneficial to the host organism.

An identical phenomenon occurs in the skin, allowing cosmetic applications to be envisaged, since a lack of equilibrium of the cutaneous microbial flora is the root of numerous cosmetic and dermatological problems. For these reasons, they are designated "GOS of interest" in the present text.

Throughout the text, polysaccharides synthesized by the glycosyltransferases of the invention are either high molecular weight dextrans when the reaction is carried out without a glucose acceptor, or oligosides when the reaction is carried out in the presence of a glucose

acceptor such as maltose or isomaltose without this necessarily being specified. The functionality of the enzyme is characterized by the nature of the glucose-glucose bonds, [α(1→6), α(1→2)] or others, and not by the molecular weight of the polysaccharide that is synthesized.

dextransucrases from L. mesenteroides already have a number of applications in industry, and in particular those from the NRRL B-1299 strain for which a method for synthesizing GOSs having α(1→2) linkages has been described in European patent EP-B1-0 325 872.

Marguerite Dols et al (2) showed that the GOS produced dextransucrases from that strain are in fact a mixture of at least three similar families of molecules differing by the number and position of the α(1→2) type linkages, which leads to the hypothesis that different glycosyltransferase type enzymatic activities exist in that bacterial strain.

Because of the industrial interest pertaining to GOSs with α(1→2) linkages as summarized above in the field of prebiotic foodstuffs, in cosmetics or in pharmaceuticals, the present invention aims to isolate and characterize a particular enzyme from among those produced by L mesenteroides NRRL B-1299 which more particularly would be involved in the synthesis of oligosides having α(1→2) linkages. The identification and characterization of such an enzyme have the advantage firstly of providing a uniform, reproducible method for producing GOSs of interest and secondly of identifying the essential characteristics of the producer enzyme for said GOSs of interest in order, if appropriate, to improve the performance of the products of the enzymatic reaction as a function of the envisaged use.

The technical problem underlying the present invention is thus to provide an enzyme and hence isolated nucleic acids encoding said enzyme to allow the improved production of GOS having α(1→2) linkages.

The present invention provides a technical solution to the various questions mentioned above by providing a novel dextransucrase, termed DSR-E, encoded by a gene endowed with a

novel and unexpected structure (dsrE) capable of catalyzing the synthesis of glucans or oligosaccharides containing α(1→2) linkages.

Between the date of filing of the priority document, French patent number 0103631 in which the dextransucrase of the invention was termed DSR-D, and that of the present application, another dextransucrase, different from the enzyme of the invention, was described and also termed DSR-D. For this reason, in the present patent application, the dextransucrase described, claimed and shown in Figure 1b) is no longer designated DSR-D as in the priority document, but is termed DSR-E. In fact, the DSR-D dextransucrases in said priority document and DSR-E are completely identical.

The term "novel and unexpected structure" means that the organization of the protein differs from that of all other glycosyltransferases described until now (1) with a catalytic domain located upstream of a glucan binding domain, the latter constituting the carboxylic portion of the protein.

The present invention thus concerns an isolated polypeptide having an enzymatic glycosyltransferase activity capable of forming dextrans having α(1→2) linkages, characterized in that it comprises at least one glucan binding domain and a catalytic activity domain located downstream of the glucan binding domain. The term "located downstream" means the fact that the amine portion of the sequence with catalytic activity or catalytic domain is proximal to the carboxylic portion of the glucan binding domain. These two domains can be immediately contiguous or, in contrast, they may be separated by a variable domain.

The glycosyltransferase of the invention preferably comprises a signal peptide.

In one implementation of the invention, the glycosyltransferase comprises two catalytic domains located either side of the glucan binding domain.

The presence of a domain with catalytic activity in the carboxylic portion of the enzyme is an essential characteristic of the latter in its capacity to form osidic α(1→2) bonds. In fact, as will be shown in the experiments described below, deletion of this domain in an enzyme having

at least two catalytic domains results in the production of glucans or oligosides essentially having α(1→6) type osidic bonds and free of α(1→2) type bonds.

More precisely, the catalytic domain, as long as it is located downstream of a glucan binding domain, allows the synthesis of oligosides containing α(1→2) bonds.

Further, the experiments described below demonstrate that the specificity of the dextransucrase DSR-E function, namely its capacity to catalyze the formation of α(1→2) osidic bonds, can be attributed not to the concomitant presence of two catalytic domains but rather to the concatenation of a glucan binding domain and a catalytic domain, and more particularly the CD2 catalytic domain.

A comparative analysis of the different glycosyltransferases including dextransucrases has demonstrated a very high degree of conservation of their catalytic domain.

The catalytic domain located in the carboxy-terminal portion of the glycosyltransferase of the invention has a sequence having at least 44% identity and 55% similarity with the catalytic domains of the other analyzed glycosyltransferases. In particular, the catalytic domain in the carboxylic portion of the glycosyltransferase of the invention has at least 65% identity and at least 80% similarity with the SEQ ID No: 1, the catalytic triad Asp/Glu/Asp in respective positions 230/268/342 being conserved.

Throughout the text, the term "X%" similarity" with respect to a reference sequence means that X% of the amino acids are identical or modified by conservative substitution as defined in the ClustalW amino acid alignment software (http:///bioweb.pasteur.fr/docs/doc-gensoft/clustalw//) and that (100-X)% can be deleted, substituted by other amino acids, or that (100-X)% can be added to the reference sequence. A particular primary structure of the enzyme of the invention is shown in SEQ ID No: 2 which represent a sequence of 2835 amino acids of a dextransucrase of L. mesenteroides NRRL B-1299.

This dextransucrase, denoted DSR-E, like most glycosyltransferases and dextransucrases, has a signal sequence, a variable domain of low conservation, a highly conserved catalytic

domain (CD1), a glucan binding domain (GBD) and a second catalytic domain (CD2) in the carboxylic portion of the protein. DSR-E is the first glycosyltransferase analyzed and has two catalytic domains, in the configuration shown in Figure 1b). It is also the first glycosyltransferase the catalytic domain of which is located in the carboxylic portion of the protein.

Figure 1b) also shows that the glucan binding domain is substantially longer than that described above for known dextransucrases; thus, a further characteristic of the enzymes of the invention is the size of this domain which is over 500 amino acids.

A comparison and analysis of the DSR-E sequence with the sequences of the glycosyltransferases or dextransucrases that have already been described (1), and the means used to this end are indicated in Example 2 detailed below. It clearly shows that while the existence of two catalytic domains substantially differentiates DSR-E from other enzymes, in contrast the sequences of said domains are substantially conserved. In particular, the amino acids necessary for catalytic activity are conserved in the second catalytic domain, namely the triad Asp/Glu/Asp located in respective positions 2210/2248/2322 of SEQ ID No: 2.

Thus, the invention also concerns any isolated polypeptide having a catalytic glycosyltransferase activity that is capable of forming dextrans or oligosaccharides having $\alpha(1\rightarrow2)$ linkages as obtained by modification, substitution, insertion or deletion of amino acid sequences but comprising sequences having at least 80% and preferably at least 90% similarity with the following sequences of SEQ ID No: 2:

423-439	2120-2138
478-501	2161-2184
519-539	2202-2214
560-571	2243-2250
631-645	2315-2322
1014-1021	2689-2696

Preferably, finally, a polypeptide with catalytic activity of the invention contains the following amino acids:

W in positions 425 and 2122;

E in positions 430, 565 and 2127, 2248;

D in positions 487, 489, 527, 638, 2170, 2172, 2210 and 2322;

H in position 637 and 2321;

Q in position 1019 and 2694.

The polypeptides with glycosyltransferase activity that can form osidic α(1→2) bonds can be in the isolated form or, in contrast, integrated into a larger protein such as a fusion protein. It may be advantageous to include sequences having another function, such as a specific tag sequence of a ligand that can facilitate purification. These tag sequences can be of the following types: GST (glutathione-S-transferase), intein-CBD (chitin-binding domain) (sold by New England Biolabs, http://www.neb.com), MBD (maltose binding domain), polypeptides containing contiguous histidine residues that can facilitate purification of the polypeptide with which it is fused. The skilled person could design any other fusion protein that could associate the function of the DSR-E of the invention with another function, a non limiting example being a sequence increasing the stability of the enzyme produced by expression in a recombinant host or a sequence that can increase the specificity or efficacy of action of said enzyme, or a sequence aimed at associating another connected enzymatic activity.

Such fusion proteins also fall within the scope of the invention provided that they contain the CD2 domain of the glucan binding site. In the same manner, fragments of SEQ ID No: 2, comprising at least SEQ ID No: 1 and the glucan binding domain, alone or integrated into a larger polypeptide forms part of the invention, as long as the enzymatic activity of the dextransucrase is conserved.

Variations of the polypeptide sequences defined above also form part of the invention. In addition to the polypeptides obtained by conservative substitution of the amino acids defined above, the variations include polypeptides the enzymatic activity of which is improved, for example by directed or random mutagenesis, by DNA shuffling, or by duplication of the CD2 catalytic domain.

The particular structure of this enzyme identified in the present invention results from a process comprising:

a) identifying and isolating dextransucrase from L mesenteroides catalyzing the production of GOSs of interest carrying α(1→2) linkages;

b) sequencing the enzyme fragments;

c) synthesizing amplification primers that can amplify the gene corresponding to the producing strain or fragments thereof;

d) sequencing the amplified fragments;

e) cloning in specific vectors and their expression in appropriate hosts.

The features of the method employed are given in detail in the experimental section below. The first step consists of separating the proteins by polyacrylamide gel electrophoresis and identifying bands having a dextransucrase activity by an in situ enzymatic reaction in the presence of substrate and acceptor. The nature of the GOSs synthesized is then identified for each band by HPLC analysis using the methods described in (1). The retention time for the oligosides in HPLC depends on the nature and organization of their osidic bonds. In particular, it is possible to distinguish between those constituted by residues having α(1→6) bonds, having α(1→6) bonds with a α(1→2) linkage at the nonreducing end of the molecule, and the desired compounds having a linear α(1→6) chain with α(1→2) linkages.

The inventors therefore isolated and identified dextransucrase from L. mesenteroides NRRL B-1299 producing GOSs of interest.

A reverse engineering process carried out in steps b) to e) above then provide the nucleotide sequence encoding the enzyme, allowing industrial scale production and, if appropriate, allowing it to be modified, improving its performance using techniques that are available to the skilled person. As an example, directed or random mutagenesis or DNA shuffling can be cited (3).

The invention also pertains to an isolated nucleic acid molecule encoding an enzyme with glycosyltransferase activity that can form dextrans or oligosides having α(1→2) linkages and comprising at least one sequence encoding a glucan binding domain, and at least one nucleotide sequence encoding a catalytic domain located on the 3' side of the foregoing, said sequence encoding a catalytic domain having at least 50% and preferably at least 70% similarity with SEQ ID No: 3.

The term "similarity" means that for the same reading frame, a given triplet is translated by the same amino acid. Thus, this term includes modifications to bases resulting in degeneracy of the genetic code.

The percentage similarity is determined by comparing a given sequence with the reference sequence. When they have different lengths, the percentage similarity is based on the percentage of nucleotides in the shortest sequence which are similar to those in the longest sequence.

The degree of similarity can be conventionally determined using software such as ClustalW (Thompson et al, Nucleic Acid Research (1994), 22: 4673-4680) distributed by Julie Thompson (Thompson@EMBL-Heidelberg.de) and Toby Gibson (Gibson@EMBL-Heidelberg.de) at the European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117, Heidelberg, Germany. ClustalW can also be downloaded from a number of websites including IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, B P 163, 67404 Illkirch Cedex, France; ftp://ftp-igbmc.u-strabg.fr/pub/) and EBI (ftp://ftp.ebi/ac.uk/pub/software/) and all sites linking to the Bioinformatics Institute (Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK).

The isolated nucleic acids of the invention can in particular comprise other sequences intended to improve the expression and/or activity of the enzyme produced.

As an example, they can be:

- sequences encoding a signal sequence for their secretion;

- duplication of the sequence encoding the CD2 catalytic domain.

Preferably, an isolated nucleic acid of the invention comprises:

a) two sequences encoding catalytic domains having at least 50%, preferably at least 80% similarity with SEQ ID No: 3;

b) a sequence enclosing the glucan binding domain, the latter preferably being located between the two sequences in a).

A nucleic acid of the invention can also comprise:

- a promoter suitable for its expression in a selected host cell;
- a sequence encoding a signal peptide; and/or
- one or more variable sequences;

said sequence or sequences all being located in the 5' portion of sequences encoding the catalytic domain(s).

A more particular example of an isolated nucleic acid of the invention comprises:

a) SEQ ID No: 4;

b) a sequence having at least 80% similarity with SEQ ID No: 4; or

c) the complementary strand to sequence a) or b); or

d) a sequence hybridizing a), b) or c).

The hybridization in d) is carried out under standard conditions, and preferably under stringent conditions. The term "hybridization under stringent conditions" means that there is at least 80% sequence identity with the sequence which is to be hybridized, preferably an identity of at least 90% of the sequence which is to be hybridized, under conditions which are, for example, described in Sambrook and Russel (3rd edition, 2001, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

The invention also concerns a gene encoding a dextransucrase that can form at least 15% $\alpha(1\rightarrow2)$ linkages. In addition to the encoding sequence, the gene comprises sequences that allow transcription initiation and sequences that allow attachment of messenger RNA to the ribosome

(RBS). SEQ ID No: 5 represents a gene structure as isolated from L. mesenteroides NRRL B-1299.

The nucleotides upstream of the translation initiation ATG are numbered 1 to 232.

The existence of a RBS sequence can be identified between nucleotides 218 and 223 as well as the consensus sequences -35 and -10 located between nucleotides 82 and 86 (TTGAA) on the one hand and 100 and 105 (ATAAAT) on the other hand.

Any nucleic acid sequence that can be hybridized with DNA of SEQ ID No: 4 or its complementary strand is capable of encoding an enzyme having the properties and characteristics of the enzyme of the invention. This applies to natural sequences existing in micro-organisms other than L. mesenteroides NRRL B-1299 and isolated from gene libraries of micro-organisms, and to those prepared by genetic engineering or by chemical synthesis.

In particular, the sequences upstream of the translation initiation ATG and necessary for expression of the protein can advantageously be substituted by transcription initiation and/or ribosome binding sequences suitable for the expression system selected for the coding sequence.

A nucleic acid sequence that is capable of hybridizing under stringent conditions with the isolated nucleic acid of the invention also comprises fragments, derivatives or allele variations of the nucleic acid sequence of the invention which encodes a protein having the enzymatic activity described above. Thus, the fragments are defined as fragments of nucleic acid molecules that are sufficiently long to encode a protein that has conserved its enzymatic activity. This also encompasses fragments that are free of the sequence encoding the signal peptide responsible for protein secretion.

The term "derivative" means a sequence that is different from the original sequence in one or more positions but which has a high degree of similarity with said sequences. In this context, "similarity" means at least 80% identity of the nucleotides, preferably at least 90% identity with the original sequence. The modifications in this case are deletions, substitutions,

insertions or recombinations provided that the enzyme encoded by these homologous sequences has the enzymatic activity of the polypeptides of the invention.

The nucleic acid sequences of the invention as described above and qualified by derivatives of said molecules as defined above are generally variations exerting the same biological function. Said variations can be natural variations, in particular those observed from one species to another and resulting in interspecies variability or, in contrast, those introduced via directed or random mutagenesis or by DNA shuffling.

Similarly, the invention encompasses isolated nucleic acids encoding a glycosyltransferase that can catalyze the synthesis of dextran or oligosaccharide carrying at least 20% and preferably at least 30% type α(1→2) linkages obtained by DNA shuffling and comprising:

- a step for random modification of one of the sequences defined above and in particular SEQ ID Nos: 3 and 4 and establishing the variations;
- a step for expressing a host housing a variation from said modified sequences in a suitable host cell,;
- a step for screening hosts expressing an enzyme that can form more than 20% and preferably more than 30% α(1→2) bonds on a suitable substrate and a step for isolating the improved gene or genes.

An isolated nucleic acid of the invention can also comprise:

a) a sequence containing at least 80% similarity with the sequence encoding a dextransucrase expressed by the plasmid pCR-T7-dsrE in E. coli deposited at the CNCM on 15th March 2001 with accession number I-2649 (E.coli JM 109 [pCR-T7-dsrD]), or

b) a complementary sequence of the sequence in a).

The denomination of the strain transformed by the recombinant plasmid pCR-T7-dsrE deposited at the CNCM is that indicated above in brackets. This does not affect the change in

the denomination of the gene carried out following deposition of said strain for the reasons given above.

The invention also concerns nucleic acid fragments as defined above, which are hybridizable with SEQ ID No: 4 and can be used as hybridization probes for detecting sequences encoding the enzymes of the invention. Said fragments can be prepared using any technique known to the skilled person.

In addition to hybridization probes, amplification primers also form part of the invention. Said primers are fragments which are hybridizable with SEQ ID No: 4 or with its complementary strand and which allow amplification of specific sequences encoding dextransucrases present in a prokaryotic or eukaryotic animal or plant organism.

The use of said amplification primers allows the use of a method for identifying the possible existence of a gene encoding an enzyme that can catalyze synthesis of GOS with $\alpha(1\rightarrow2)$ linkages in said organism, said method also forming part of the invention.

The invention also concerns expression vectors comprising a nucleic acid as described above under the control of a sequence allowing its expression and preferably its excretion in prokaryotic or eukaryotic cells. The term "prokaryotic cells" preferably denotes bacteria selected from a group comprising E. coli, Lactococcus, Bacillus and Leuconostoc. The term "eukaryotic cells" preferably means eukaryotes selected from a group containing yeasts, fungi and plants.

The vector comprises a promoter suitable for expression of the isolated nucleic acid of the invention in the selected expression system. As an example, the T7 bacteriophage promoter could advantageously be selected for expression in E.coli.

The invention also concerns host cells, prokaryotic or eukaryotic, transformed by a nucleic acid of the invention, preferably comprised in an expression vector carrying a promoter, adapted for expression in the selected host cells. The transformed cells are selected from Gram-

bacteria such as E.coli, or from Gram+ bacteria such as Lactococcus, Bacillus, Leuconostoc, or from eukaryotes in a group comprising yeasts, fungi and plants.

One particular example of a cell transformed in accordance with the invention is the E. coli strain harboring a plasmid termed pCR-T7-dsrE carrying the SEQ ID No: 4 under the control of the T7 bacteriophage promoter deposited at the CNCM on 15th March 2001 under accession number I-2649.

The present invention also concerns a method for producing a glycosyltransferase that can form dextrans or oligosides having at least 15% and preferably at least 20% of type α(1→2) osidic linkages and comprising:

a) inserting a nucleic acid or a vector as defined above into a host cell that can express and preferably secrete the glycosyltransferase;

b) characterizing the enzymatic activity being investigated using any of the methods accessible to the skilled person;

c) purifying the enzyme from a cell extract.

The term "method for characterizing enzymatic activity known to the skilled person" means the methods described in the literature, for example in reference (2), and novel methods that may be developed to allow identification and discrimination of glucooligosaccharides having the desired degree of linkages.

In fact, it concerns any screening method that can identify the presence of α(1→2) linkages in a GOS.

Examples are:

- HPLC in which GOS migration varies as a function of the nature and position of the linkages, in particular those containing the α(1→2) bond at the reducing end and those containing this bond on the penultimate glucose; and/or
- nuclear magnetic resonance (NMR);

- the existence of a positive reaction with specific monoclonal antibodies of α(1→2) bonds on the reducing end and/or specific monoclonal antibodies of α(1→2) bonds on the penultimate glucose of the GOS.

The invention also concerns a method for obtaining a glycosyltransferase that can have oligosides or dextrans having a percentage of α(1→2) linkages of more than 15% and preferably more than 30% of the totality of the osidic bonds and comprising a step for modifying SEQ ID No: 4 by addition, deletion or mutation provided that:

- the reading frame is not modified; and
- the following amino acids are conserved after translation:

W in positions 425 or 2122, encoded by the TGG triplet in positions 1273 and 6364;

E in positions 430, 565, 2127 and 2248, encoded by GAA triplets in positions 1288, 1693, 6379 and 6742 respectively;

D in positions 487, 489, 527, 638, 2170 and 2210, encoded by GAT triplets in positions 1459, 1465, 1579, 1912, 6508 and 6628 respectively;

D in positions 2172 and 2322 encoded by GAT triplets in positions 6514 and 6964;

H in position 637 and 2321, respectively encoded by the CAT triplet in position 1909 and CAC in position 6961;

Q in positions 1019 and 2694, respectively encoded by triplets CAA (position 3055) and CAG (position 8080).

A method for producing a glycosyltransferase according to the invention having the same characteristics as above can also comprise:

- a step for randomly modifying SEQ ID No: 4 and establishing a library of variations;
- a step for expressing a host housing a variation from said modified sequences in a suitable host cell,;

- a step for screening hosts expressing an enzyme that can form more than 15% and preferably more than 30% of α(1→2) bonds on a suitable substrate;
- and a step for isolating the improved gene or genes.

In a further implementation of the invention, the method consists of modifying SEQ ID No: 3 by duplicating all or part of the CD2 catalytic domain.

It should be understood that the methods above are not only aimed at obtaining a glycosyltransferase that can form oligosides having a constant and reproducible percentage of α(1→2) linkages of more than 15% of the total linkages, but also to improve the degree of α(1→2) linkages with the aim of modifying the properties of the oligosides obtained to improve their dietetic properties or their capacity to maintain or re-establish bacterial flora associated with certain organs of the human or animal body.

Finally, the present invention concerns glycosyltransferases that can be obtained by a method as defined above and which can form at least 15% and preferably at least 30% of type α(1→2) osidic linkages in glucooligosaccharides.

Finally, the invention pertains to the use of glycosyltransferases of the invention as well as those that can be obtained by the methods mentioned above, in the production of a composition with a pre-biotic effect or in the manufacture of a dermatological, cosmetic or pharmaceutical composition.

Non-limiting examples that can be cited are the improvement in intestinal transit in animals and in man, an improvement in calcium and/or magnesium assimilation and of minerals in general, preventing cancer of the colon and prevention or treatment of skin affections such as acne, dandruff or body odor.

The advantage of the polypeptides and nucleic acids encoding said polypeptides of the invention is not only in improvements in terms of quality, yield, reproducibility and cost of glycosyltransferases that can form oligosaccharides with type α(1→2) osidic linkages, but also in producing novel enzymes the functionality of which is improved.

The figures, examples and detailed description below provide non-limiting illustrations of the particular characteristics and functionalities of polypeptides with enzymatic activity and sequences encoding them. In particular, they can illustrate more precisely the specificity of the catalytic domain present in the carboxylic portion of the enzyme of the invention and its potential evolution to obtain improved enzymes.

KEY TO FIGURES

Figure 1: Structure of native glycosyltransferases and derived recombinant proteins: Figure 1a) shows the structure of glycosyltransferases and dextransucrases described in the literature (1). PS: signal peptide; ZV: variable zone; CD: catalytic domain; GBD: glucan binding domain. Figure 1b) shows the structure of the glycosyltransferase of the invention. Figures 1c) to 1i) show different constructions comprising deletions in comparison with native DSR-E protein. Δ(PS) corresponds to the control constituted by the entire form cloned into the pBAD-TOPO thiofusion system (Invitrogen).

Figure 2: Diagrammatic summary of the method for cloning the nucleotide sequence encoding a glycosyltransferase of the invention using a genome library by using a PCR probe described in Table I and a HindIII/EcoRV probe respectively.

Figure 3: Comparison of the signal sequences of different glycosyltransferases of L. mesenteroides. The conserved amino acids are shown in bold. DSR-B: L. mesenteroides NRRL B-1299 (4); DSR-S: L. mesenteroides NRRL B-512F (5); ASR: L. mesenteroides NRRL B-1355 (6).

Figure 4: Alignment of 11 repeat sequences of the DSR-E enzyme and observed in the variable zone.

Figure 5: Alignment of conserved sequences in the catalytic domain.

- Block A: essential amino acids of the N-terminal portion of the catalytic domain;
- Block B: amino acids of the catalytic saccharose binding domain;

- Blocks C, D, E: blocks containing three amino acid residues involved in the catalytic triad (6);
- Block F: sequence containing glutamine 937 of GTF-1 studied by Monchois et al (7).

The entirely conserved amino acids are indicated in bold. "*": conservative substitutions; ":": semi-conservative substitutions; ---: gap. The numbering is that for SEQ ID No: 2.

Figure 6 : HPLC characterization of products synthesized by recombinant enzyme DSR-E.

6A: HPLC analysis of glucooligosaccharides obtained with dextransucrases of L. mesenteroides NRRL B-1299.

6B: HPLC analysis of glucooligosaccharides obtained by recombinant DSR-E. The following peaks are identified:

1: fructose
2: maltose;
3: sucrose;
4: panose;
5: R4;
6: OD4;
7: R5;
8: OD5;
A, B, C: unidentified peaks.

6C: recombinant DSR-E deleted from the catalytic domain of the carboxylic portion of the enzyme (ΔDSR-E).

Figure 7: HPLC analysis of acceptor on maltose reaction products synthesized by different entire forms and deleted from the DSR-E protein.

L.m. B-1299: mixture of dextransucrases produced by L. mesenteroides NRRL B-1299.

The peaks were identified as follows:

F: fructose;

M: maltose;

S : saccharose

P: panose;

R4, R5: GOS comprising α(1→2) bonds;

OD4, OD5: GOS free of α(1→2) bonds.

MATERIALS AND METHODS

1) Bacterial strains, plasmids and growth conditions:

All strains were kept at -80°C in tubes containing 15% glycerol (v/v).

Leuconostoc mesenteroides B-1299 (NRRL, Peoria, USA) was cultivated at 27°C with stirring (200 rpm) on standard medium (saccharose 40 g/l, potassium phosphate 20 g/l, yeast extract 20 g/l, $MgSO_4,7H_2O$ 0.2 g/l, $MnSO_4,H_2O$ 0.01 g/l, NaCl 0.01 g/l, $CaCl_2$ 0.02 g/l, $FeSO_4,7H_2O$ 0.01 g/l), the pH being adjusted to 6.9.

Escherichia coli DH5α and JM109 were cultivated on LB medium (Luria-Bertani).

Selection of pUC18 or pGEM-T Easy recombinant clones was carried out on LB-agar dishes supplemented with 100 µg/ml of ampicillin, 0.5 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) and 40 µg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). E. coli TOP 10 cells were used to clone the PCR TOPO Cloning (Invitrogen) product and cultivated on LB medium supplemented with kanamycin in a concentration of 50 µg/ml.

Regarding expression of dsrE, the ECHO Cloning System cloning kit (Invitrogen) allows a PCR product to be cloned in a donor vector (pUNI/V5-His-TOPO), preceding a step for recombination with a suitable acceptor vector (pCR-T7-E). This system requires E. coli PYR1, TOP 10 and PL21(DE3)pLysS cells cultivated on LB medium supplemented with 50 µg/ml of kanamycin as well as 34 µg/ml of chloramphenicol for the BL21(DE3)pLysS strain.

Digested and dephosphorylated pUC18 plasmids from Pharmacia (Amersham Pharmacia Biotech) were used to constitute the genomic DNA library of L. mesenteroides NRRL B-1299. Cloning of the PCR product necessitated the use of the pGEM-T Easy plasmid (Promega) and TOPO-XL plasmid (Invitrogen) for fragments of more than 2 kbp.

The pBAD-TOPO Thiofusion system (Invitrogen) used to construct the different deleted forms of the DSR-E protein used the araBAD promoter the control mechanisms for which

involve the AraC regulatory protein. In the absence of an inducer, namely L-arabinose, the dimeric AraC protein associates with the regulatory structures of the operon and entrains the formation of a DNA loop, said loop then blocking transcription of genes placed under the control of the araBAD promoter. In the presence of L-arabinose, in contrast, AraC forms a complex which liberates the DNA loop and allows transcription initiation. The base expression can be limited by adding glucose to the culture medium: this reduces the level of cyclic AMP and thus concomitant activation of the CAP protein (cAMP activator protein). The level of activation obtained is a function of the concentration of L-arabinose so that the optimum conditions for production of the protein of interest can be selected with accuracy.

Further, the use of this vector can allow a 12 kDa thioredoxin tag to be positioned on the N-terminal end of the protein of interest. This fusion encourages the translation of the gene encoding said protein of interest. The tag protein can also enhance the solubility of the protein to which it is fused. The pBAD-TOPO Thiofusion system is designed to allow ready elimination of the thioredoxin tag by simple cleavage using enterokinase. Finally, using this expression system, a histidine tag is inserted on the C-terminal end side of the protein of interest. Such a tag is used to purify said protein by affinity.

Within the context of using this system, the E. coli TOP 10 strain was cultivated on LB medium supplemented with 100 µg/ml of ampicillin.

2) Gel electrophoresis, location and characterization of enzyme:

After culturing L. mesenteroides NRRL B-1299 for 7 h, the medium was centrifuged (7000 rpm, 4°C, 30 min) and the cells, in which 90% of the enzymatic activity was found, were concentrated 10 times in an acetate buffer solution (20 mM, pH 5.4), heated for 5 minutes at 95°C in the presence of denaturing solution (tris HCl 62.5 mM, SDS 4%, urea 6M, bromophenol blue 0.01% and β-mercaptoethanol 200 mM). 300 µl of the mixture was deposited on 7% polyacrylamide gel. After migration, the total proteins were revealed by amido black staining, while the dextransucrase activity was detected by staining with Schiff's reagent polymer after

synthesizing the dextran in situ. The bands corresponding to the active dextransucrases were excised and incubated separately in 2 ml of 20 mM sodium acetate solution, pH 5.4, containing 100 g/l of saccharose and 50 g/l of maltose. After total consumption of saccharose, the reaction was stopped by heating to 95°C for 5 minutes, and the reaction medium was centrifuged for 5 minutes at 15000g to eliminate the insoluble dextran. The samples were analyzed by reverse phase chromatography (C18 column, Ultrasep 100, 6 µm, 5x300 mm, Bishoff Chromatography) using ultrapure water as the eluent, at a constant flow rate of 0.5 ml/min. The oligosaccharides were separated for 30 minutes at ambient temperature and detected by refractometry. Peptide sequencing was carried out on the selected protein bands by the Laboratoire de Microséquençage, Institut Pasteur, Paris.

3) Molecular biological techniques used

Purification of the E. coli plasmid and purification of the genomic DNA of L. mesenteroides was carried out using the QiaPrep Spin Plasmid kit and the Cell Culture DNA maxi kit (Qiagen) respectively. The amplification and cloning methods were carried out using standard techniques (Sambrook and Russel, 2001, supra). Restriction and modification enzymes from New England Biolabs or Gibco BRL were used in accordance with the manufacturer's instructions.

PCR was carried out with primers selected on the basis of the protein sequence obtained on an isolated band from gel electrophoresis (see supra, gel electrophoresis and enzyme localization). Two peptides were selected:

- 29-FYFESGK; and
- 24-FESQNNNP

and used to synthesis degenerate oligonucleotides indicated in Table I below.

In this table, the numbering of which follows that of SEQ ID No: 4, it appears that the presence of a serine residue in the two peptides necessitates the synthesis of two primers for each peptide since serine can be encoded by six different codons. ECHO-dir and ECHO-inv are

primers which allowed amplification of dsrE by PCR for cloning into the ECHO Cloning (Invitrogen) expression system.

TABLE I

Désignation	Description	Séquence 5'-3'
29-dir1	FYFESGK	TT(C/T)TA(C/T)TT(C/T)GA(A/G)*TCA*GG(C/G)AA(A/G)
29-dir2		TT(C/T)TA(C/T)TT(C/T)GA(A/G)*AGC*GG(C/G)AA(A/G)
24-inv1	FESQNNNP	(T/G)GG(G/A)TT(G/A)TT(G/A)TTTTG*TGA*(T/C)TCAAA
24-inv2		(T/G)GG(G/A)TT(G/A)TT(G/A)TTTTG*GC*T(T/C)TCAAA
IPCR-rev	séquence nt 5769-5798	CCCTTTACAAGCTGATTTTGCTTATCTGCG
IPCR-dir	séquence nt 8311-8342	GGGTCAAATCCTTACTATACATTGTCACACGG
ECHO-dir	séquence nt -6 - 39	AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG
ECHO-inv	séquence nt 8457-8504	ATTTGAGGTAATGTTGATTTATCACCATCAAGCTTGAAATATTGACC

PCR

PCR was carried out using a Perkin-Elmer thermocycler, model 2400, with 50 nanograms of genomic DNA. The quantities of primers used was 10 µM of 29-Dir-1 and of 24-Inv1. 250 µM of each triphosphate deoxynucleotide and Taq polymerase were added to the reaction mixture.

After amplification of 25 cycles at 94°C for 30 seconds then at 50°C for 30 seconds, then at 72°C for 5 minutes, a 666 base pair fragment was obtained.

Certain fragments were amplified using the "Expand Long Template PCR" (Roche Boehringer Mannheim) system, in accordance with the manufacturer's instructions. This system can amplify large fragments of up to about 20 kbp highly effectively. The combination of two DNA polymerases can minimize errors during the elongation phases.

Southern hybridization and gene library of L. mesenteroides NRRL B-1299

Chromosomal DNA from L. mesenteroides NRRL B-1299 was digested with different restriction enzymes then separated by electrophoresis on 0.8% agarose gel in TAE 0.5X buffer.

Genomic libraries of the bacteria were transferred onto nylon hybond N+ membranes (Amersham PharmiciaBiotech). Hybridization was carried out using the 666 base pair fragment of deoxy-adenosine-triphosphate labeled with ^{32}P. The labeling reaction was carried out using the "Mega Prime DNA Labelling System Kit" (Amersham PharmaciaBiotech) labeling kit, followed by purification of the probe on MicroSpin S-200HR columns. Pre-hybridization and hybridization was carried out under highly stringent conditions (65°C overnight using the normal methods) (Sambrook and Russel, 2001, supra).

Reverse PCR

The reverse PCR reaction produces a linear DNA fragment from a circular matrix using divergent primers.

Genomic DNA from L. mesenteroides NRRL B-1299 was digested with EcoRV under the conditions recommended by the manufacturer.

After re-circularization, the digestion products were used as a matrix in a reverse PCR reaction [Extrapol II DNA polymerase (Eurobio), reaction volume of 50 µl, reverse PCR reaction parameters: 25 cycles; 94°C; 30 seconds; 51°C, 30 seconds; 72°C, 3 minutes]. The two primers were selected as a function of the pSB2 insert sequence as indicated in Figure 2.

Figure 2 summarizes the conditions for obtaining different plasmids carrying dsrE fragments by screening the gene library and using the probes described above.

DNA sequence and analysis

After sequencing the peptides, degenerate primers marked out because of the frequency of use of codons in the dextransucrase genes of L. mesenteroides NRRL B-1299 were synthesized and allowed amplification of a 666 bp fragment. Sequencing this fragment revealed strong homologies with the genes of known dextransucrases, even though it was entirely novel.

The use of this fragment as a homologous probe in Southern experiments allowed positive signals on different tracks of genomic DNA to be marked. A first HindIII library was then screened and a recombinant plasmid termed pSB2 containing a 5.6 kbp insert was purified.

An analysis of the sequence for this HindIII fragment revealed an open reading frame covering the whole insert. Then a EcoRV library was screened with a HindIII/EcoRV probe isolated at the N-terminal end of the 5.6 kbp HindIII insert. A recombinant pSB3 recombinant plasmid, tested positively by dot-blot, was shown to contain a 3.8 kbp insert which, after sequencing, was shown to contain the initiation codon for translation and the promoter region of the novel dextransucrase gene termed dsrE.

With the aim of obtaining the dsrE termination codon, reverse PCR was carried out on genomic DNA from L. mesenteroides NRRL B-1299 digested with EcoRV and re-ligated to itself, using divergent oligonucleotide primers designated from the pSB2 insert sequence. A single fragment with the expected size of 1 kbp was amplified then cloned in pGEM-T Easy to obtain the pSB4 plasmid. After sequencing, the amplified sequence located downstream of the HindIII site comprised 221 bp and contained the reading frame termination codon for dsrE located 30 bp downstream of the HindIII restriction site.

Sequencing of the different fragments carried by the three plasmids was carried out on both strands by the company Genome Express. Sequence analyses of the nucleotides was carried out using "ORF Finder" (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), Blast (http://www.ncbi.nlm.nih.gov/blast/blast.cqi, Altschul et al, 1997), ClustalW (http://www2.ebi.ac.uk/clustalw, Thompson et al, 1994), PRODOM (http://protein/tolulouse.inra.fr/prodom.html, Corpet et al, 2000), PFAM (http://pfam.wustl.edu.hmmsearch.shtml, Bateman et al, 2000) and SAPS (http://bioweb.pasteur.fr/segana/interfaces/saps.html, Brendel et al, 1992), all of this software being available on the Internet.

Protein expression

Two cloning and expression systems were used to produce recombinant proteins in E. coli, namely the ECHO-Cloning and pBAD-TOPO Thiofusion (Invitrogen) systems.

By way of example, the method for cloning the nucleotide sequence encoding the DSR-E protein using the ECHO-Cloning system will now be briefly described.

Two primers as proposed in Table I above were used for amplification using the "Expand Long Template" system under the following conditions: 94°C for 3 minutes, followed by 25 cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 68°C for 7 minutes. The PCR products were then cloned into the pUNI/V5-His-TOPO vector to obtain a donor vector (pUNI-dsrE) to be recombined with an acceptor vector (pCR-T7-E) and adapted for expression in E. coli. The final plasmid was designated pCR-T7-dsrE.

This construction, placing the dsrE gene under the control of the bacteriophage promoter T7, allowed inducible expression of the dsrE gene.

After induction with 1 mM of IPTG, the transformed E. coli BL21 cells were harvested by centrifuging after 4 hours growth and re-suspending at a final optical density of 80 at 600 nm in a 20 mM sodium acetate buffer, pH 5.4, and 1% Triton X100 (v/v) in the presence of 1 mM of PMSF to prevent proteolysis in the cell extracts after sonication.

Similar experiments carried out with the pBAD-TOPO Thiofusion system allowed the recombinant vector pBAD-TOPO-dsrE to be constructed.

Enzymatic tests

The enzymatic reactions were carried out under standard conditions at 30°C in a 20 mM sodium acetate buffer, pH 5.4, NaN_3 1 g/l and saccharose, 100 g/l. The activity of the DSR-E enzyme was determined by measuring the rate at which the reducing sugars were liberated, represented here by fructose, using the dinitrosalicylic acid method which is well known to the skilled person. One unit is defined as the quantity of enzyme which would catalyze the formation of 1 μmol of fructose per minute under standard conditions. The oligosaccharides were synthesized in a reaction medium containing 100 g/l of maltose, 200 g/l of saccharose and 0.5 units/ml of DSR-E.

As for the dextran synthesis, the enzymatic reaction was continued for 24 hours in the presence of 100 g/l of glucose. The dextran produced was precipitated in the presence of 50% (v/v) ethanol and washed twice in 50% ethanol (v/v) prior to freeze drying. It was then dissolved in an amount of 10 mg/ml in D_2O and analyzed by ^{13}C NMR spectrometry.

HPLC separation

100 µl samples were removed and heated at 95°C for 5 minutes then diluted in ultrapure water to obtain a final concentration of total sugars of less than 5 g/l. After centrifuging, the residual substrates and the different species formed were analyzed by HPLC on a C18 column (Ultrasep 100, 6 µm, 5x300 mm, Bishoff Chromatography).

The oligosides were separated at ambient temperature for 30 minutes in ultrapure water used as the eluent, at a flow rate of 0.5 ml/min. Detection was accompanied by refractometry.

These conditions allowed the following species to be separated: fructose, maltose, leucrose, saccharose, and oligosides with a degree of polymerization that did not exceed 6.

Calculation of yields

The method for calculating the yields for the oligoside synthesis reactions took into account the residual concentration of the acceptor in accordance with the following formula:

R = {[GOS final] – [initial GOS]}/{0.474 x [sacchraose consumed] + [acceptor consumed]}

in which R represents the real yield of the total GOS synthesis reaction, the concentrations being expressed in g/l.

Construction of different deleted forms of DSR-E protein

The different deleted forms of the DSR-E protein [Figure 1c) to 1i)] were obtained by PCR amplification of fragments corresponding to the dsrE gene then cloning in the pBAD-TOPO Thiofusion vector described above. The primers used for amplification of the regions selected from the dsrE gene are shown in Table II below. The positions of the primers are shown

with respect to SEQ ID No: 5, relating to the sequence for the dsrE gene. The bases mutated to introduce the NcoI restriction site are shown in bold and the resulting NcoI site is underlined.

TABLE II

Désignation	Positions	Séquence 5'-3'
pBAD-PS/ZV-dir	344-373	G**CCATGG**CAAATACGATTGCAGTTGACACG
pBAD-ZV/CD1-dir	971-1001	G**CCATGG**ACGGTAAAACCTATTTTCTTGACG
pBAD-CD1/GBD-dir	3656-3682	T**CCATGG**GTGAAAAAACAAGCACCGGC
pBAD-GBD/CD2-dir	6167-6189	A**CCATGG**ATATGTCTACTAATGC
pBAD-CD1/GBD-inv	3638-3658	TAACTGTTTAGGCAAGAATCC
pBAD-GBD/CD2-inv	6146-6172	TAATGTATTAGTGAATAAGTATTCACC
pBAD-ent-inv	8714-8737	AATTTGAGGTAATGTTGATTTATC

The above direct and reverse primers were designed to ensure translational fusion of the N-terminal thioredoxin tag and the C-terminal polyhistidine tag of the truncated protein forms, satisfying the open reading frames for the regions encoding said forms.

If the pBAD-TOPO Thiofusion plasmid contains a specific restriction site for the NcoI enzyme located at the 5' end of the region encoding thioredoxin, a second NcoI site is introduced into each direct primer to enable extraction of that region if required.

The PCR amplification reactions were carried out using the "Expand Long Template" system under the following conditions: pre-denaturing at 94°C for 3 minutes followed by 25 cycles at 94°C for 30 seconds, 52°C for 30 seconds and 68°C for 7 minutes.

The amplification products generated were then cloned into the pBAD-TOPO Thiofusion vector for subsequent transformation of the E. coli TOP 10 strain. Recombinant clones were selected, their restriction profile analyzed to identify a recombinant plasmid carrying the insertion orientated as expected for each of the investigated forms.

Example 1: Characterization and purification of the DSR-E enzyme and obtaining the dsrE gene.

The enzymes produced by L. mesenteroides cultures and obtained on a polyacrylamide gel in SDS as described in the Materials and Methods section were isolated by cutting the gel.

The GOSs produced by the isolated enzymes were analyzed by HPLC using the methods described in (1). The enzyme the activity of which was being investigated was deduced from the nature of the GOSs produced. After trypsic proteolysis and separation of the peptides produced by HPLC, 2 peptides: 29-FYFESGK and 24-FESQNNNP, were sequenced and used as a model for the synthesis of degenerate nucleotide primers.

The different amplification and cloning steps are shown in Figure 2. The complete gene was inserted into the pCR-T7-E plasmid and expressed in E. coli.

The production of a functional enzyme was attested by the production of GOSs the HPLC analysis of which is shown in Figure 6b).

The size of peaks 5 and 7, representing GOSs with a $\alpha(1\rightarrow2)$ linkage, should in particular be noted.

Example 2: Characterization of dsrE and DSR-E sequences

2.1 Nucleotide sequence

The nucleotide sequence of the enzyme is shown in SEQ ID No: 4. It is composed of a reading frame of 8506 nucleotides.

The nucleotide sequence for insertion into the pCR-T7-dsrE plasmid contained a ribosome binding site (RBS), 9 bases upstream of the ATG initiation codon and was composed of a hexanucleotide GAGGAA.

2.2 Analysis of amino acid sequence

The 8506 nucleotide dsrE sequence encodes a 2835 amino acid protein shown in SEQ ID No: 2. The isoelectric point for this protein is 4.88 and its theoretical molecular weight is 313.2 kDa. Despite strong similarities with known dextransucrases, DSR-E is characterized by an original structure.

Alignment of the amino acid sequence with known glycosyltransferases and dextransucrases confirmed that the structure in the glycosyltransferase domain and dextransucrases domain was conserved, namely: a signal sequence, a variable zone, a highly conserved catalytic domain and a glucan binding domain. This structure is shown in Figure 1a).

As indicated in Figure 1b), a second catalytic domain forms the carboxy-terminal portion of the enzyme, as confirmed by PRODOM and Blast analysis.

With a molecular weight of 313.2 kDa, DSR-E had about twice the mean molecular weight of other glycosyltransferases and dextransucrases (1), which is in agreement with the presence of a second catalytic domain at the c-terminal end and also with a longer glucan binding region.

a) Analysis of signal sequence:

The signal sequence and the nucleotide sequence encoding the peptide signal were highly conserved if compared with other dextransucrases, as shown in Figure 3. The cleavage site is located between amino acids 40 and 41.

b) Variable domain:

Downstream of the signal peptide, DSR-E had a 207 amino acid variable domain. When it was compared with other variable glycosyltransferase domains, using a SAPS type alignment program, the presence of a 14 amino acid motif repeated 11 times was revealed, as indicated in Figure 4.

This alanine-, threonine- and aspartic acid-rich repeat motif has never before been identified.

The role and significance of this region has never been elucidated. Different studies have shown that its deletion does not affect enzymatic activity (4). The role of the 14 amino acid repeat motif, which does not exist in other glycosyltransferases, remains to be determined, however.

c) Analysis of catalytic domains:

The first catalytic domain extends from amino acids 248 to 1142 (CD1) of SEQ ID No: 2, while the second is located between amino acids 1980 and 2836 (CD2). These two domains have 45% identity and 65% similarity between them.

CD1 and CD2 contain amino acids already identified in glycosyltransferases and dextransucrases as being essential to their enzymatic activity, as shown in Figure 5.

The catalytic triads of CD1 and CD2 determined by analogy with α amylase (7) are present in the following positions:

(Asp 527/Glu 565/Asp 638 for CD1 and Asp 2210/Glu 2248/Asp 2322 for CD2).

Other conserved residues were identified as being important for enzymatic activity: the residues Trp 425/Glu 430 for CD1 and Trp 2122/Glu 2127 for CD2, which are analogous to those of the N-terminal domain of GFT1 described by Monchois et al (4): Trp 344/Glu 349.

In contrast, certain sequences located in the conserved region of the glycosyltransferases and dextransucrases are not found in the CD2 of DSR-E. Thus, as indicated in Figure 5 below, the sequences FIHNDT (2214-2220) and KGVQEKV (2323-2329) diverge from other consensus sequences of dextransucrases already studied, which are respectively NVDADLL and SEVQTVI.

d) Glucan binding domain:

When the DSR-E sequence is compared with known sequences, it appears that the glucan binding region is substantially longer. In fact, the length of this domain is about 500 amino acids in the glycosyltransferases and dextransucrases being studied while in DSR-E, it represents 836 amino acids. Several A and C repeat motifs, in particular a series of AC repetitions, have been identified. Table III below shows the consensus sequences of the repeat motifs of GBD, in particular the A and C motifs, described in the literature relating to dextransucrases of Leuconostoc and Streptococcus spp.

TABLE III

Motif	Consensus sequence
A	WWYFNxDGQAATGLQTIDGQTVFDDNGxQVKG

B	VNGKTYYFGSDGTAQTQANPKGQTFKDGSGVLRFYNLEGQYVSGSGWY
C	DGKIYFFDPDSGEVVKNRFV
D	GGVVKNADGTYSKY
N	YYFxAxQGxxxL

x: any amino acid

EXAMPLE 3: Expression of dsrE in E. coli

E. coli BL21 (DE3) pLysS pCR-T7-dsrE cells were cultivated as described above. After polyacrylamide gel electrophoresis (page-SDS), analysis of the protein extracts effectively revealed the presence of several bands having saccharase dextran activity, said activity being measured as described above.

The E. coli JM109 [pCR-T7-dsrD] line was deposited at the CNCM on 15th March 2001 with accession number I-2649.

Identification and characterization of enzymatic activity

Using a glucose acceptor molecule, the dextransucrases produced by recombinant E. coli were compared with those produced by L. mesenteroides NRRL B-1299.

HPLC analysis of the reaction products with recombinant DSR-E (Figure 6) showed retention times corresponding to the previously identified GOSs R4 and R5 (2). Type R oligosaccharides are linear GOS series, the α(1→2) bond being linked to the non-reducing end. The OD series, linear GOSs resulting from glycoside α(1→6) bonds with a maltose residue at the reducing end was observed in very small quantities. Three novel compounds, in contrast, were detected in the recombinant enzyme products.

Identification of GOSs produced:

Finally, Figure 6b clearly shows that peaks 5 and 7 representing the GOSs of the R series are relatively larger with the recombinant enzyme than with the native enzyme in which the peaks corresponding to panose and OD5 are larger.

Example 4: Effect of deletion of CD2 on the enzymatic activity of DSR-E

The genomic DNA of L. mesenteroides NRRL B-1299 was used as a matrix to amplify the dsrE gene by PCR deleted from the sequence corresponding to the second catalytic domain.

To this end, 2 oligonucleotides, ECHO-dir (5'-AGTTGTATGAGAGACATGAGGGTAATTTGTGACCGTAAAAAATTG) (SEQ ID No: 6) corresponding to the nucleotide sequence -6 to 39 and containing the translation initiation codon, and ECHO-inv-del (5'-GTATTAGTGAATAAGTATTCACCATTGCATTTATCGTCAAAATAGTACG) (SEQ ID No: 7) complementary to the sequence 5889-5937 and corresponding to the peptide sequence YYFDDKGNGEYCFTNT, were synthesized, to fuse the C-terminal end of the deleted protein with a His tag present on the cloning vector. The PCR reaction was carried out using a DNA thermal cycler model 2400 (Perkin Elmer) with the Expand Long Template System (Boehringer Mannheim) using the following temperature cycle: 94°C for 3 min, then 25 cycles with: 30 s at 94°C, 30s at 55°C and 7 min at 68°C. The PCR product was then cloned into the pUNI donor vector and the resulting plasmid was used in a recombination reaction with the pCR-T7-ΔdsrE expression vector.

The cell extract, preparation, enzymatic reaction and reaction product analysis were those described in Example 3 above.

The HPLC profile of the GOSs obtained with the DSR-E enzyme deleted from the CD2 domain appear in Figure 6c).

The type R GOS shown as peaks 5 and 7 shown in Figure 6a) and 6b) are entirely absent from the products obtained with the recombinant enzyme deleted from CD2. The only analyzable products were those corresponding to linear oligosides resulting from α(1→6) bonds with a maltose residue in the reducing portion. This result clearly indicates the essential role of the catalytic domain located in the carboxy-terminal portion of the enzyme in its capacity to form α(1→2) osidic bonds.

Example 5: Study of structure-function relationships of DSR-E protein

The dsrE gene, insofar as it is the first gene encoding a dextransucrase catalyzing the synthesis of α(1→2) bonds to have been cloned, is of particular interest. Thus, it is important to

characterize this gene and its expression product, in this case by determining the roles of the different domains making up the DSR-E protein in the function which has been assigned thereto, namely to correspond to a α(1→2 specific to the synthesis of α(1→2) bonds.

5.1 Deleted forms of DSR-E protein:

A study of six different forms obtained by deletion of one or more domains from the DSR-E protein was envisaged in order to determine the following by reference to Figure 1 below: (i) the influence of the presence of the CD2 domain by studying GBD-CD2 and Δ (CD2) constructions; (ii) the influence of the presence of the variable zone by analyzing the Δ(ZV) and CD1-GBD forms; and (iii) the intrinsic catalytic potential of the CD1 and CD2 domains expressed in an isolated manner (CD1 and CD2 constructions).

The catalytic activity of each of the different forms was compared with that observed with the control corresponding to the entire form deleted from the single signal peptide Δ(PS) [Figure 1c)].

5.2 Analysis of constructions:

At the end of the experimental PCR amplification and cloning procedure detailed above, several clones with an insertion in the expected orientation were obtained for each of the envisaged constructions, with the exception of the truncated GBD-CD2 form for which the desired amplification product could not be cloned.

The sequences for the insertions were determined in order to ensure the absence of mutations that after translation may modify the amino acids located at positions presumed essential for the enzymatic activity of the protein encoded this way.

A mutation was identified at the 31st insertion base relative to the control Δ(PS), inducing substitution of one aspartic acid by an asparagine in position 10 of the variable zone. As it is not located in the repeat motifs S of the variable zone (Figure 4), it appears that the incidence of this mutation on the finally observed function is negligible.

A mutation was introduced into the amplification product corresponding to the construction Δ(CD2), modifying the aromatic residue F1411 in leucine. This mutation was located in the first third of the glucan binding domain GBD at a junction between two repeat motifs.

Because of the errors made by polymerase during PCR amplification, the construction Δ(ZV) did not have the expected sequence. In fact, the insertion contained an open reading frame, that frame essentially corresponding to the GBD-CD2 form which could not be cloned. However, in the GBD-CD2 form obtained definitively in place of Δ(ZV), 46 N-terminal residues were absent. Now, the GBD domain has more than 800 amino acids forming a concatenation of 24 repeat units. This concatenation is such that, over the 46 truncated residues, only the last 9 were located at one of said units, in particular at the first thereof. It thus appears plausible to consider that deletion of these amino acids has no influence on the enzymatic reaction catalyzed by the corresponding protein form. This hypothesis supported by the fact that in other dextransucrases, the loss of a certain number of repeat units from the GBD domain does not significantly reduce the activity of the resulting protein.

The insertion encoding the CD1-GBD form contained a mutation affecting the F633 residue located in the CD1 domain and more precisely in the region that is highly conserved in dextransucrases, itself located just in front of the second aspartic acid of the catalytic triad (Figure 5). The expected phenylalanine was substituted by a leucine. It is difficult at this stage to estimate the impact of such a mutation on the observed catalytic activity.

The sequence of insertions encoding the catalytic domains CD1 and CD2 was determined in the same manner as for the other constructions.

5.3 Expression products and enzymatic activities

The proteins corresponding to the various deleted forms of DSR-E were expressed by subjecting the recombinant E. coli cells to induction by L-arabinose in a concentration of 0.002%. The enzymatic activity was observed for the first four hours following induction.

The protein extracts obtained by sonication of the cell residues were analyzed by SDS-PAGE electrophoresis (Sambrook and Russel, 2001, supra). The molecular masses of the recombinant proteins were estimated from the electrophoretic profiles obtained, said masses essentially corresponding to the expected masses taking into account the 12 kDa incrementation linked to the thioredoxin tag. Table IV below summarizes the estimated values for the molecular masses of the different truncated forms and, by way of comparison, provides the expected masses.

TABLE IV

Protein form	Expected mass (kDa)	Expected mass + thioredoxin (kDa)	Estimated mass (kDa)
Δ(PS)	309	321	324
Δ(CD2)	218	230	ND
GBD-CD2	224	/	233
CD1-GBD	193	205	199
CD1	99	111	111
CD2	95	107	ND

ND: not determined

Table V below indicates the nature and position of amino acids marking the start and end of the protein forms constructed in this study. The different positions refer to SEQ ID No: 2 corresponding to the protein DSR-E.

TABLE V:

Protein form	Starting amino acid	Ending amino acid	Total length
Δ(PS)	N41	I2835	2795
Δ(CD2)	M1	L1980	1980
GBD-CD2	M1188	I2835	1648
CD1-GBD	I248	L1980	1733
CD1	I248	Q1141	894
CD2	D1981	I2835	855

The GBD-CD2 form did not have a thioredoxin tag. In fact, this form was derived from experimental uncertainty occasioned by the procedure for PCR amplification of the sequence assumed to encode the Δ(ZV) form. Because of the deletions from the sequences thus generated, the thioredoxin tag, in principle situated at 5' from the protein of interest, could not be fused with the GBD-CD2 region.

The quality of the electrophoresis gels did not allow determination as to whether the level of expression of the different forms was quantitatively identical and as a result whether said forms were present in the same proportions in the cell extracts.

The activity measurements provided were established on the basis of a given volume of cell extracts but could not be extrapolated to the quantity of each protein actually contained in said volume of extracts.

The synthesis of dextran polymers in situ by incubating electrophoresis gels in a saccharose solution and subsequent staining with Schiff's reagent confirmed the presence of proteins having a glucan-saccharase activity in cell extracts corresponding to Δ(PS), Δ(CD2), GBD-CD2 and CD1-GBD.

Table VI below shows the maximum enzymatic activities observed for each construction. The results confirm the data drawn from the experiments in which the gels were stained with Schiff's reagent, namely the fact that the cell extracts relative to the forms Δ(PS), Δ(CD2), GBD-CD2 and CD1-GBD had a saccharase activity, in contrast to the two catalytic domains taken in isolation. This result was in agreement with the literature, given that it has been demonstrated that in other dextransucrases, the absence of the GBD domain induced a drastic loss of enzymatic activity (8, 9, 10).

TABLE VI

Protein form	Δ(PS)	Δ(CD2)	GBD-CD2	CD1-GBD	CD1	CD2
maximum activity (U/I)	1063	181	86	235	5.3	0

The intrinsic activity of the CD1 form was too low to be detected. Regarding the GBD-CD2 form, it had a non negligible activity which leads to the conclusion that the corresponding structural organization, namely a catalytic domain downstream of the glucan binding domain, remains enzymatically active.

5.4 Effect of deletions on oligoside synthesis:

Provided that the specificity of the synthesis of α(1→2) bonds was conserved during the reaction in the presence of an acceptor, experiments for synthesizing oligosides starting from maltose were carried out (Figure 7).

When the reactions were carried out to completion, i.e. all of the saccharose had been consumed, the oligoside synthesis yields were calculated. The results are shown in Table VI below. Only the reaction involving the cell extract containing the protein form CD1 did not allow such a calculation. The temperature effect probably resulted in inactivation of the very low activity present in the protein extract.

TABLE VII

Protein form	Yield of oligosides in OD series (%)	Yield of oligosides in R series (%)	Total oligoside yield
Native enzyme	36	28	64
Δ(PS)	41	14	55
Δ(CD2)	67	1	68
GBD-CD2	45	47	92
CD1-GBD	100	0	100

As indicated in Figure 7 below, the presence of oligosides from series R was only detected with enzymatic forms having the catalytic domain CD2, with the exception of the case in which said domain was isolated and then rendered completely inactive. In fact, the retention time for the oligosides synthesized by the deleted form of the second catalytic domain and by the CD1-GBD form corresponded only to those in the OD series, i.e. to GOSs deprived of α(1→2) bonds. These results thus indicate that the CD2 domain was required for the formation of α(1→2) bonds.

The products obtained with the GBD-CD2 form have supported these observations. This construction, which had CD2 as the only catalytic domain, was capable of catalyzing in a preponderant manner the synthesis of oligosides from the R series, having α(1→2) bonds. Thus, this results demonstrates that specificity in terms of the function of the DSR-E enzyme resides in the highly original sequence for this domain, and not in the association of two catalytic domains. Further, the GBD-CD2 protein form also allowed the synthesis of α(1→6) bonds. However, the low yields obtained for these oligosides indicated that they were preferentially converted into

oligosides with a higher degree of polymerization belonging to the R series, which prevented their accumulation in the reaction medium, differing from molecules from the R series which were not converted (2).

By comparing the profiles of the products obtained as shown in Figure 7, it is clear that the entire form Δ(PS) mainly synthesizes linear oligosides. In fact, the molecule R4 was absent and the oligoside R5 only present in a small amount. The catalytic domain CD1 catalyzed the exclusive synthesis of α(1→6) bonds and its activity appeared to be preponderant with respect to that of the CD2 domain. In addition, in the entire form of the enzyme, the implication of the CD2 domain would thus be less important because of: (i) lower intrinsic catalytic parameters; and/or (ii) a global enzyme configuration that was unfavorable to its activity.

Further, the entire enzyme Δ(PS) catalyzed the synthesis of oligosides from the R series with a lower yield than that observed with the mixture of dextransucrases produced by L. mesenteroides NRRL B-1299 (Figure 7). The yield obtained, 28%, was situated between those observed for the entire form Δ(PS) and for the GBD-CD2 form. It is known that the wild strain produces several forms of dextransucrases that are susceptible of synthesizing osidic bonds, in particular α(1→2) bonds. One hypothesis has been proposed, in which said forms are the degradation products of DSR-E. Insofar as the truncated forms of DSR-E such as GBD-CD2 could catalyze the synthesis of oligosides from the R series more effectively, it would appear that the yields obtained with the heterogeneous mixture produced by L. mesenteroides NRRL B-1299 can be attributed to the contribution of the catalytic activities of the ensemble of said different enzymatic forms.

In conclusion, by isolating a particular dextransucrase produced by L. mesenteroides, the inventors have succeeded in characterizing a particular and unexpected structure of this enzyme that can produce oligosides of interest and have α(1→2) type linkages. Identification and characterization of this sequence allows the construction of recombinant cells or organisms

specifically expressing this enzyme and also allows its modification by directed or random mutagenesis or by DNA shuffling to further improve its characteristics to be envisaged.

This invention can also improve the yield and reproducibility of the production of GOSs of interest for the different applications cited above.

REFERENCES

(1) Monchois V., Willemot R.M., Monsan P. (1999). Glucansucrases: mechanism of action and structure-function relationships. FEMS microbiol. Rev. 23,131-151.

.(2) Dols M., Remaud-Simeon M., Willemot R.M., Vignon M.R., Monsan P.F. (1998). Structural characterization of the maltose acceptor-products synthesised by Leuconostoc mesenteroides NRRL B-1299 dextransucrase. Carbohydrate Research 305, 549-559.

(3) Arnold F.H. (2001). Nature 409 n° 6817, 253.

(4) Monchois V. Vignon M., Russel R.R.B. (1999). Isolation of key amino-acid residues at the N-terminal end of the core region of Streptococcus downei glucansucrase GTF-I. Appl. Microbiol. Biotechnol. 52, 660-665.

(5) Wilke-Douglas M., Perchorowicz J.T., Houck C.M., . 20 Thomas B.R. (1989). Methods and compositions for altering physical characteristics of fruit and fruit products. PCT patent, WO 89/12386.

(6) Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarçabal P., Willemot R.M., Monsan P. (2000). Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355. FEMS Microb. Lett. 182, 81-85.

(7) Devulapalle K.S., Goodman S., Gao Q, Hemsley A., Mooser G. (1997). Knowledge-based model of a glusocyltransferase from oral bacterial group of mutant streptococci. Protein Sci. 6, 2489-2493.

(8) Kato C., and Kuramitsu H.K. (1990). Carboxy-terminal deletion analysis of the Streptococcus mutans glucosyl-transferase-1 enzyme. FEMS Microbiol. Lett. 72, 299-302.

(9) Lis M., Shiroza T., et Kuramitsu H.K. (1995). Role of the C- terminal direct repeating units of the Streptococcus mutans glucosyltransferase-S in glucan binding. Appl. Env. Microbiol. 61, 2040- 2042.

(10) Monchois V., Remaud-Simeon M., Russel R.R.B., Monsan P. and Willemot R.M. (1997). Characterization of Leuconostoc mesenteroides NRRL B-512F dextransucrase (DSR-S) and identification of amino-acid residues playing a key role in enzyme activity. Appl. Microbiol. Biotechnol. 48, 465-472.